Shahrood University of Technology

Journal of Hydraulic and Water Engineering (JHWE)

Journal homepage: https://jhwe.shahroodut.ac.ir

Iranian Hydraulic Association (IHA)

Comparison of the Performance of PSO and GA Algorithms in Predictive Modeling of Flood-Related Deaths in Boma

André Mampuya Nzita [1,2*], Clément N'zau Umba-di-Mbudi [2,3], Fils Makanzu Imwangana [3,4], Guyh Dituba Ngoma [5]

[1] Regional School of Water (ERE), University of Kinshasa (UNIKIN), Kinshasa, Democratic Republic of the Congo
[2] President Joseph Kasa-Vubu University, Faculty of Engineering, Boma, Democratic Republic of the Congo
[3] University of Kinshasa, Faculty of Science and Technology, Kinshasa, Democratic Republic of the Congo
[4] Center for Research of Geological and Mining, Unit of Geomorphology and Remote Sensing, Kinshasa, P.Box. 898 Kinshasa I, DR. Congo
[5] University of Quebec in Abitibi-Temiscamingue, School of Engineering, Rouyn-Noranda, Canada.

Article Info

Article history:
Received 15 May 2025
Revised form 30 June 2025
Accepted 10 August 2025
Published online 16 August 2025

DOI:
10.22044/JHWE.2025.16251.1064

Keywords
Floods
Predictive modeling
River dynamics,
PSO and GA optimization algorithms,
Boma

Abstract

This study examines river dynamics and flooding in the town of Boma, Democratic Republic of Congo, where vulnerability to flooding is increased by climate change and anthropogenic pressures. This study aims to address gaps in flood-related fatality prediction by developing a predictive model incorporating the interaction between the Congo River water level and the Kalamu River discharge. The objectives include using a generalized linear model (GLM) with a Poisson distribution, combined with optimization algorithms such as particle swarm optimization (PSO) and genetic algorithms (GA). The methodology relies on the collection of historical data on water levels, discharges, rainfall, and fatalities, followed by rigorous data analysis using preprocessing and optimization techniques. The results show that PSO outperforms GA in terms of convergence speed and efficiency, achieving a better fitness value. Fitness values reveal an RMSE of 8.37, an MAE of 6.42, and an R^2 of -4.04, indicating significant inaccuracies in the forecasts. Simulations reveal a direct relationship between water level, discharge, and deaths, highlighting the importance of these interactions for risk management. These results provide valuable tools for infrastructure planning and raising awareness of the impact of floods on vulnerable populations, thus contributing to more effective prevention strategies.

1. Introduction

River dynamics and flooding pose a growing threat to riverine populations, particularly in vulnerable regions (Jian et al., 2021), such as Boma, a town located along the Congo River in the western part of the Democratic Republic of Congo and crossed by the Kalamu River. Boma's geographical location makes it particularly susceptible to flooding, a risk exacerbated by climate change and anthropogenic pressures on river ecosystems. This vulnerability is all the more worrying because it is further compounded by climate change and anthropogenic pressures on river ecosystems. Extreme events, such as floods, are becoming increasingly

• *Corresponding author:* simmlicem57@gmail.com, Tel:+243999340214

frequent, endangering populations living in riparian areas. According to the technical services of Congolaise des Voies Maritimes (CVM) (the Congolese Maritime Authority), "If this rate of rising waters of the Congo River continues, there would be an overflow of the Kalamu River and its tributaries beyond their banks," recalling the devastating events of 2016 that put the town of Boma at risk. Indeed, a critical gap persists in accurately predicting flood-related deaths, particularly in combining hydrological data, advanced statistical modeling, and optimization techniques (Tien Bui et al., 2016). Recent scientific literature highlights the growing importance of predictive modeling in disaster management, as demonstrated by the study by Qi et al. (2021), which highlights the applications of urban flood models in mitigation strategies (Fernández & Lutz, 2010; Hoang & Liou, 2024; Jian et al., 2021; Zhao et al., 2019). However, many existing models face limitations, including their inability to fully capture nonlinear interactions between key variables such as water level and discharge, which compromises the reliability of risk assessments.

To address this imperative, this study aims to develop a robust and innovative predictive model for flood-related deaths in Boma. The originality of this approach lies in its ability to explicitly account for the complex interaction between the Congo River water level and the Kalamu River discharge, relying on a combination of advanced methods. The research objectives are threefold: to statistically model the relationship between water level, discharge, and flood-related deaths using a generalized linear model (GLM) with a Poisson distribution, leveraging the flexibility and power of this statistical approach to capture nonlinear relationships; to optimize the GLM parameters using a hybrid approach combining simplified particle swarm optimization (PSO) and genetic algorithms (GA), to efficiently explore the parameter space and identify optimal configurations; and to simulate flood scenarios and visualize the predicted impact on mortality rates using 3D surface plots, to facilitate communication of results and support decision-making. Previous research, such as Khosravi et al. (2019) and Teng et al. (2017), demonstrates the importance of optimization in flood risk modeling, highlighting the effectiveness of optimization algorithms in this context (Barbulescu, 2025; Chapi et al., 2017; Deng et al., 2022; Jahandideh-Tehrani et al., 2020; Kalantar et al., 2021; Tien Bui et al., 2016; Tuyen et al., 2021; Yu et al., 2023; Zhen & Bărbulescu, 2025). To achieve these ambitious goals, comprehensive historical data on Congo River water levels (1960-2017), Kalamu River discharges, rainfall, evaporation (1992-2023), and flood-related deaths were collected from reliable sources, such as the CVM in Boma, the Boma weather station, and local flood documentation. These data underwent rigorous preprocessing, including scaling of water-level and discharge variables, to ensure comparability and suitability for statistical models. The PSO and GA algorithms were carefully implemented and calibrated to optimize model parameters, minimizing an objective function defined as the negative log-likelihood of the Poisson distribution plus an L1 regularization term to avoid overfitting and ensure robustness. Previous studies, such as Lui et al. (2023), Mudashiru et al. (2021), and Fernández and Lutz (2010), highlight the importance of data preparation and model optimization in flood forecasting.

Finally, the optimized model was used to simulate realistic flood scenarios, and the predicted mortality rates were visualized using 3D surface plots, enabling intuitive interpretation of the results. This integrated and rigorous approach will improve our understanding of the complex flood dynamics and their impact on human lives in the Boma region and provide valuable tools for risk management and decision-making. The theoretical framework underlying this research draws on several key areas: hydrological modeling (Hassan et al., 2021), statistical regression (Noor et al., 2022; Rima et al., 2025; Nguyen, 2020), and optimization algorithms. Hydrological modeling provides a basis for understanding the relationships among rainfall, river discharge, and water levels. Statistical regression, particularly GLM with a Poisson distribution, offers a powerful tool for modeling count data, such as the number of flood-related deaths, while accounting for multiple predictor variables. Optimization algorithms, such as PSO and GA, provide efficient methods for finding the best set of model parameters to minimize prediction errors. Moreover, model regularization is crucial for avoiding overfitting and ensuring generalization to new data. Previous research, such as that by Di Baldassarre et al. (2013) and Rentschler et al. (2022), highlights the importance of these concepts for developing effective predictive models.

2. Materials and Methods

2.1. Presentation of the study environment

The study was conducted in the town of Boma and in the Kalamu River catchment, as shown in Figures 1 and 2. Before flowing into the Congo River near the port of Boma, the Kalamu River flows through three communes of the town: Kabondo, Nzadi, and Kalamu. The catchment of the Kalamu River covers an area of 68.84 km² with a perimeter of 44.4 km.

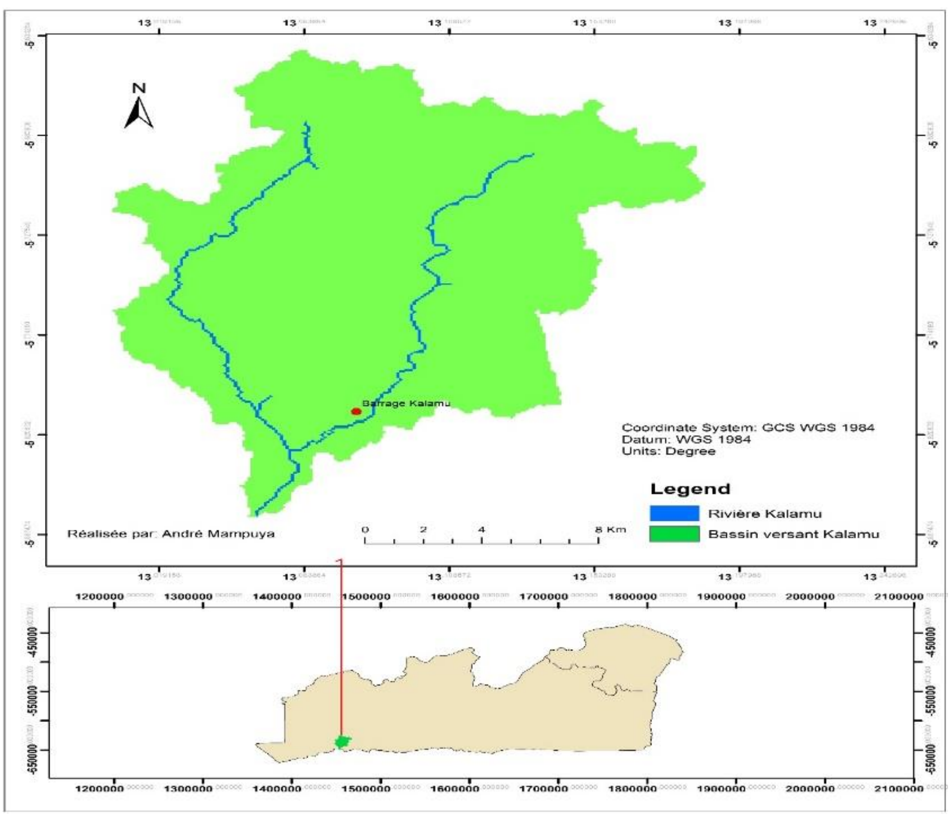

Figure 1. Catchment of the Kalamu River at Boma

2.2. Data collection

Crucial data were collected from various sources (Table 1). The CVM in Boma recorded water levels of the Congo River from 1960 to 2017. Monthly precipitation and evaporation data were obtained from the Boma meteorological station, covering the period from 1992 to 2023. Historical information on Kalamu River flows and the number of deaths in years corresponding to floods was provided by flood-related documentation in Boma.

Table 1 illustrates the data for the flood study variables.

Table 1. Variables associated with flooding in Boma by year

Year of floods	Water level of the Congo River (m)	Congo River flood	Discharge of the Kalamu River on flood day (m^3/s)	Number of deaths from flooding	Flood day precipitation (mm)
23/12/1985	2.6	Large	20.53	3	207.9
12/12/1999	3.39	Large	36.32	0	365.8
30/12/2000	2.83	Large	25.85	0	260.6
20/12/2010	2.79	Large	24.97	0	252.1
12/12/2015	3.31	Large	24.51	4	247.1
26/12/2016	3.5	Large	47.62	40	458.6
05/12/2018	3.01	Large	15.53	0	158.1
26/11/2019	3.24	Large	12.71	2	129.7
03/11/2021	2.93	Large	23.84	2	240.9
03/12/2022	2.74	Large	23.76	0	240

2.3. Combination of optimization algorithms (PSO and GA), statistical modeling (Poisson GLM), and simulation techniques

The analysis begins with data preparation. Historical data from the Python Anaconda software, including year, water level, discharge, number of deaths, and precipitation, is loaded into a Pandas DataFrame. To ensure optimal performance of the optimization algorithms, a feature scaling step is performed on the Water_Level and Flow_Rate variables using the StandardScaler class from scikit-learn (Chapi et al., 2017). This transformation prevents variables with a wide range of values from exerting disproportionate influence on the optimization process. The scaling is performed according to the following equation:

$$X\_Scaled = \left(\frac{X - mean(X)}{std(X)}\right) \quad (1)$$

Where X represents the original variable, mean (X) is its mean, and std (X) is its standard deviation (Deng et al., 2022).

Next, an interaction term, Flow_Water_Interaction, is created by multiplying the scaled values of Water_Level and Flow_Rate. This new variable is intended to capture potential nonlinear relationships between water level and discharge and their combined impact on the number of deaths. The calculation of this interaction term is defined by:

$$Flow_{Water_{Interaction}} = Scaled\_Water\_Level \times Scaled\_Flow\_Rate \quad (2)$$

At the heart of the optimization process is the objective function, objective_function. This function quantifies how well a given set of parameters (a particle's position in PSO or an individual's genes in GA) fits the observed data. The goal of both PSO and

GA algorithms is to minimize this function. The objective function is based on a Poisson regression model, an appropriate choice given that the deaths variable represents count data (non-negative integers). Poisson regression models the occurrence rate of events (deaths in this case) (Jahandideh-Tehrani et al., 2020). More precisely, the objective function is based on the negative log-likelihood of the Poisson distribution. Minimizing the negative log-likelihood is equivalent to maximizing the likelihood of observing the actual data, given the model. The Poisson distribution is defined by:

$$P(Y = y) = \frac{\mu^{y} \times e^{(-\mu)}}{y!} \tag{3}$$

Where y is the observed number of deaths and μ Is the expected number of deaths per unit time (the rate). The log-likelihood is then:

$$\log L(\mu) = \sum \begin{pmatrix} y_i \times log(\mu_i) \\ -\mu_i - \log(y_i!) \end{pmatrix} \tag{4}$$

Where y_i is the observed number of deaths for observation i and μ_i is the expected number of deaths for observation i. The negative log-likelihood is simply the opposite of this value:

$$-\log L(\mu) = -\sum \begin{pmatrix} y_i \times log(\mu_i) - \\ \mu_i - log(y_i!) \end{pmatrix} \tag{5}$$

The expected value μ is modeled as an exponential function of the input variables and their coefficients:

μ= exp (intercept + Water_Level_coeff × Water_Level + Flow_Rate_coeff × Flow_Rate + interaction_coeff × Flow_Water_Interaction) (6)

To avoid overfitting, an L1 regularization term (Lasso) is added to the objective function. L1 regularization adds the sum of the absolute values of the coefficients to the loss function, which encourages model parsimony by reducing some coefficients to zero. The regularization term is defined by:

$$\text{Regularization} = \lambda_{reg} \times \sum\left(abs(x)\right) \tag{7}$$

Where λ_{reg} is the regularization strength and x represents the coefficients.

The combined objective function is therefore:

$$\text{Objective} = -\log \mathrm{L}(\mu) + \text{Regularization} \tag{8}$$

The Simplified Particle Swarm Optimization (PSO) algorithm is a population-based optimization algorithm inspired by the social behavior of flocks of birds or schools of fish. A swarm of particles explores the solution space, with each particle adjusting its position based on its own best-known position and the best-known position of the entire swarm (Khosravi et al., 2019). Initially, the particle positions (X) and velocities (V) are randomly initialized. In addition, the best-known position of each particle (Pbest) is initialized to its initial position, and the most prominent position of the entire swarm (Gbest) is determined.

At each iteration, each particle updates its velocity based on three factors: its previous velocity (inertia), a cognitive component (attraction to its own best position), and a social component (attraction to the swarm's best position). The equation governs the velocity update:

V[i] = inertia_weight × V[i] + cognitive_coefficient×rand()× (Pbest[i] - X[i]) + social coefficient ×rand()× (Gbest - X[i]) (9)

Where inertia_weight controls the influence of the particle's previous velocity, cognitive_coefficient the influence of its own best position, social_coefficient the influence of the swarm's best position, and rand() generates a random number between 0 and 1.

The position of each particle is then updated by adding its velocity to its current position:

$$X[i] = X[i] + V[i] \quad (10)$$

The objective function is evaluated for each particle's new position. If a particle's new position has a better fitness value than its previous best position, Pbest is updated. Similarly, if a particle's new position has a better fitness value than the swarm's best position, Gbest is updated. The PSO algorithm finally returns Gbest (the best solution found), its fitness value, and the history of fitness values. The Genetic Algorithm (GA) is another population-based optimization algorithm inspired by natural selection. A population of candidate solutions (individuals) evolves through processes of selection, crossover (recombination), and mutation. Initially, a population of candidate solutions is initialized with random values (Teng et al., 2017). At each iteration, individuals are selected for breeding based on their fitness, using a tournament selection method. Selected individuals (parents)
exchange genetic material to create new offspring through a single-point crossover. If the parents are P1 = [a, b, c, d] and P2 = [e, f, g, h] and the crossover point is 2, then the offspring would be O1 = [a, b, g, h] and O2 = [e, f, c, d].
Random modifications are introduced into the offspring's genomes to maintain diversity and explore new regions of the solution space. The mutation is performed according to the equation:

$$\text{offspring}[i][j] = \text{offspring}[i][j] + \text{random\_noise} \quad (11)$$

The objective function is evaluated for each offspring. The offspring then replace the worst-performing individuals in the population, ensuring gradual population improvement over time. The GA algorithm returns the best solution found, its fitness value, and the history of fitness values (Babatunde et al., 2015). After PSO optimization, the code uses the statsmodels library to fit a generalized linear model (GLM) with a Poisson family. This provides a more standard statistical framework for analyzing the relationship between predictors and the response variable. The GLM uses a Poisson distribution to model the Deaths variable, which is appropriate for count data. The log link function relates the linear predictor to the expected value of the Poisson distribution. The smf.glm function is used to fit the GLM. The formula argument specifies the model structure, and the data argument provides the data. Crucially, the start_params argument uses the coefficients obtained from the SPSO optimization as starting values for the GLM fitting process. This can help the GLM converge faster and more reliably.
The glm_model.summary() method displays a summary of the GLM results, including coefficient estimates, standard errors, p-values, and goodness-of-fit statistics. The glm_model.pseudo_rsquared method computes McFadden's R-squared, a pseudo-R-squared for GLMs. It indicates the proportion of the response variable's variance that the model explains. The code displays the fitted GLM equation, showing the estimated coefficients for each predictor.
Finally, the code includes simulation and visualization functions. The simulate_deaths function simulates the number of deaths for given water_level and flow_rate values, using the optimized parameters (PSO's Gbest or GA's best_solution). The input features (water_level and flow_rate) must be scaled using the same StandardScaler object that was used to scale the training data. This ensures that the simulation is performed at the same scale as the model was trained on. The function calculates the expected value mu using the GLM equation and then samples a value from a

Poisson distribution with mean mu. This simulates the random variation in the number of deaths. np.vectorize is used to efficiently apply the simulate_deaths function to a grid of water_level and flow_rate values. The code uses matplotlib to create 3D surface plots showing the simulated number of deaths as a function of water_level and flow_rate. This allows you to visualize the model's predictions and how the number of fatalities varies depending on different combinations of input characteristics.

This code performs a comprehensive analysis of the relationships among water levels, flows, and deaths, using a combination of optimization algorithms (PSO and GA), statistical modeling (Poisson GLM), and simulation techniques. The code also includes visualizations to help understand the model's behavior and predictions.

3. Results and Discussions

3.1. Comparative analysis of convergence and performance of PSO and GA algorithms

Figure 2 shows the performance of two optimization algorithms, PSO and GA, over iterations. The y-axis represents the Best Fitness (Negative Log-Likelihood + Regularization), and the x-axis represents the number of Iterations.

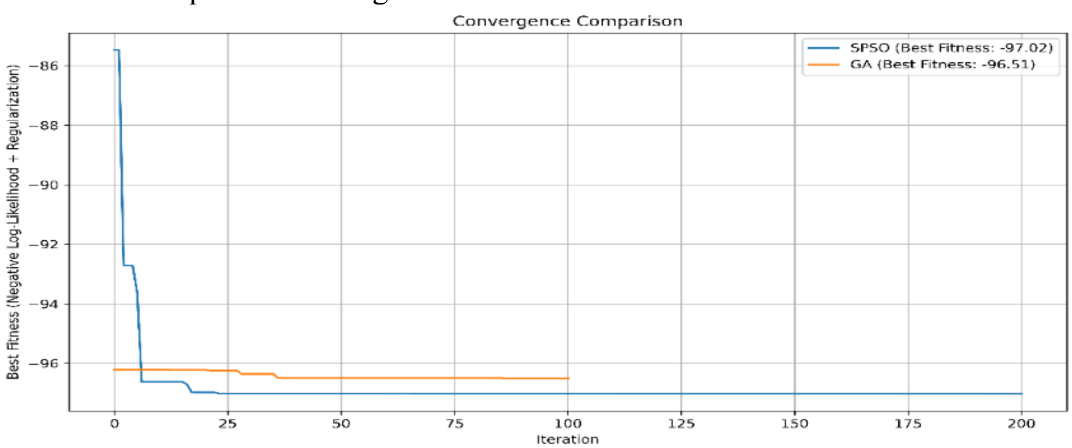

Figure 2. Convergence Comparison

In terms of global convergence, both algorithms, PSO (Synchronous Particle Swarm Optimization) and GA (Genetic Algorithm), aim to minimize the best fitness value (Negative Log-Likelihood + Regularization). A lower best fitness indicates better performance. Both algorithms exhibit a decreasing trend in the best-fitness value as the number of iterations increases, indicating convergence towards an optimal solution. Regarding the PSO algorithm, we observe a rapid improvement at the beginning. The PSO exhibits a rapid decrease in best fitness during the first few iterations (approximately iterations 0 to 25). This suggests that the PSO quickly explores the search space and quickly finds promising solutions. The PSO convergence value converges to a best fitness of approximately -97.02. The PSO convergence speed appears to converge relatively soon, with most of the improvements occurring during the first 25 iterations. After that, the change in best fitness is minimal. In terms of final performance, the PSO achieves a better best fitness value (-97.02) than the GA (-96.51), indicating that it finds a slightly better solution in this particular run. Regarding the GA algorithm, we observe a slower initial improvement. The GA shows a slower initial decrease in best fitness compared to the PSO. This suggests that the GA can take more iterations to explore the search space and find good solutions. The GA's convergence value converges to a best fitness of approximately -96.51. The GA's convergence speed converges more slowly than the PSO. The improvement in best fitness is more gradual over the iterations. In terms of final performance, the GA's final best fitness value (-96.51) is slightly worse than that of the PSO (-97.02), indicating that it did not find as optimal a solution as the PSO in this run. In terms of convergence speed, the PSO converges faster than the GA, achieving most of its improvements in the first few iterations. In terms of final performance, the PSO achieves a slightly higher best fitness value than the GA, suggesting a better final solution. In terms of algorithm behavior, PSO appears to be more efficient at quickly finding a good solution, while GA explores the search space more gradually.

In this convergence comparison, PSO appears to be the best-performing algorithm for this specific problem. It converges faster and achieves a slightly higher best fitness value than GA. However, it is essential to note that these results are specific to this particular run and problem. Algorithm performance may vary depending on the problem's characteristics and the parameters set by the algorithms.

3.2. Statistical modeling and simulation of flood-related deaths: Comparison of algorithms

In this analysis, we examine the performance of the model used to predict flood-related deaths in Boma, based on cross-validation fitness values, as well as the results of the generalized linear model (GLM) presented in Table 2.

First, the fitness values are concerning. The root mean square error (RMSE) is 8.37, meaning that, on average, the model's predictions deviate from the actual values by 8.37 units. This high value indicates significant inaccuracies in the model, which is particularly problematic for flood fatality predictions. Furthermore, the mean absolute error (MAE) is 6.42, which reinforces the idea of unsatisfactory model performance, as it indicates an average prediction error of 6.42 units. Regarding the coefficient of determination R^2, its value is -4.04. A negative R^2 is alarming because it suggests that the model cannot explain the data's variance better than a simple mean. This suggests a poor fit of the model to the data.

Table 2 provides general information about the fitted model, including the dependent variable, number of observations, model family, link function, log-likelihood, deviance, and pseudo R-squared.

Table 2. Results of the Stats models Generalized Linear Model (GLM)

Generalized Linear Model Regression Results			
Dep. Variable	Deaths	No. Observations	10
Model	GLM	Df Residuals	6
Model Family	Poisson	Df Model	3
Link Function	Log	Scale	1.000
Method	IRLS	Log-Likelihood	-19.269
Date:	Sun, 11 May 2025	Deviance	21.522
Time	21:04:57	Pearson chi2	17.5
No. Iterations:	6	Pseudo R-squ. (CS):	1.000

The model used is a generalized linear model (GLM) with a Poisson distribution and a logarithmic link function. This is appropriate because the number of deaths is a discrete, non-negative variable, which fits well with a Poisson distribution. The logarithmic link function transforms a linear combination of predictors into an occurrence rate (the expected number of deaths). To understand the statistical model's overall performance, several key elements must be examined. First, the variable being explained, or predicted, is the number of deaths. The model was built using 10 observations. It is important to note that this relatively small number of observations can potentially make the results less reliable and more sensitive to extreme or atypical values. Next, we consider the residual degrees of freedom, which amount to 6. This figure shows the difference between the total number of observations and the number of parameters the model must estimate, including the intercept and the coefficients for the various explanatory variables. Furthermore, the model uses 3 predictors: water depth, river discharge, and their interaction. The model scale is 1.0000. In the context of a Poisson model, this value is ideal if the model is correctly specified. The log-likelihood, which measures how well the model fits the data, is -19.269. The higher (i.e., less negative) this value, the better the model fit. The deviance, which quantifies the difference between the current model and a perfect (so-called saturated) model, is 21.522. A lower deviance indicates that the model is closer to an ideal fit to the data. Pearson's chi-square, another measure of fit, is 17.5. Cameron and Windmeijer's pseudo R-squared, which attempts to quantify the proportion of variance explained by the model, is 1.000. Although this may seem excellent, such a high value should be interpreted with caution, as it may indicate overfitting, particularly given the small number of observations. The estimation algorithm required six iterations to converge on a solution. Finally, the covariance type used is non-robust, meaning that the standard errors of the estimated coefficients are not protected against potential violations of model assumptions, such as heteroscedasticity.

Table 3 presents the estimated coefficients for each variable in the model, along with their standard errors, z-statistics, p-values (P>|z|), and confidence intervals.

Table 3. Regression Coefficients of the Generalized Linear Model (GLM)

Covariance Type	nonrobust					
	coef	std err	z	P>|z	[0.025	0.975]
Intercept	-0.422	0.464	-0.910	0.363	-1.330	0.487
Water_Level	0.695	0.359	1.938	0.053	-0.008	1.398
Discharge	-1.273	0.646	-1.972	0.049	-2.538	-0.008
Flow_Water _Interaction	1.595	0.446	3.575	0.000	0.721	2.470

To fully understand how the water level and discharge of the Kalamu River influence the number of deaths during floods, it is essential to consider the coefficients of the variables in the statistical model. Because the model is logarithmic, these coefficients do not translate directly into linear changes in the number of deaths but rather into multiplicative changes in the death rate. First, the model's intercept is -0.4218. Taking the exponential of this value ($\exp(-0.4218) \approx 0.656$), we obtain an estimate of the expected death rate when all other variables are zero. However, this value must be interpreted with caution, as it may not be realistic or meaningful to assume that the water level and discharge are simultaneously zero. Second, the coefficient for water level is 0.6952. This means that for every unit increase in water depth, the death rate is approximately doubled (exp(0.6952) ≈ 2.004). Thus, an increase in water depth is associated with a significant increase in the number of deaths. It should be noted that the p-value for this coefficient is 0.053, indicating marginal significance at the 0.05 level. The coefficient for river discharge is -1.2731. Taking the exponential of this value (exp(-1.2731) ≈ 0.280), we find that for every unit increase in discharge, the death rate is multiplied by approximately 0.280. This suggests that increased discharge is associated with a significant decrease in mortality. The p-value of 0.049 for this coefficient is significant at the 0.05 level. Finally, the interaction term between discharge and water depth is substantial. Its coefficient is 1.5954, which means that the effect of water depth on the number of deaths depends on discharge, and vice versa. More precisely, for each unit increase in the interaction, the death rate is multiplied by approximately 4.929 (exp(1.5954) ≈ 4.929). The p-value of 0.000 associated with this interaction term is highly significant, confirming its importance in the model.

These coefficients suggest that water depth, when considered alone, tends to increase the number of deaths during floods. Conversely, river discharge, when considered alone, tends to decrease the number of deaths. However, the interaction between these two variables is crucial, as it indicates that the effect of one on the number of deaths depends on the value of the other. For example, high discharges could mitigate the negative impact of high water levels by allowing water to drain more quickly, thereby reducing flood risks. Conversely, high water levels combined with low discharges could exacerbate risks and lead to more deaths.

The equation for the Generalized Linear Model (GLM) simulated using the SPSO and GA algorithms to predict the number of deaths is as follows:

$$Deaths = \exp\begin{pmatrix} -0.42 + 0.7 \times Wata_{Lever} - 1.27 \times \\ Flow_{Rate} + 11.60 \times \\ Flow\_Water\_Interaction \end{pmatrix} \quad (12)$$

Where Water Level represents the water level, Discharge represents the discharge, and Flow Water Interaction represents the interaction between water level and discharge. The McFadden R-squared for this model is 0.999, indicating an excellent fit to the data.

The model equation, Deaths = exp(-0.42 + 0.70 × Water_Level + -1.27 × Flow_Rate + 1.60 × Flow_Water_Interaction), provides a mathematical representation of the relationship between the variables. This equation expresses the expected number of deaths as a function of water level (Water_Level), river discharge (Flow_Rate), and their interaction (Flow_Water_Interaction). Specifically, the equation indicates that the number of deaths equals the exponential of a linear combination of the explanatory variables. The term -0.42 represents the intercept, which is a constant. The term 0.7 × Water_Level suggests that an increase in water level is associated with an increase in deaths (since the coefficient is positive). The term -1.27 × Flow_Rate indicates that an increase in discharge is related to a decrease in the number of deaths (since the coefficient is negative). Finally, the term 1.60 × Flow_Water_Interaction captures the combined effect of water level and discharge, and its positive coefficient indicates that their interaction increases the number of deaths. It is important to note that, due to the exponential function, the effects of the variables are not linear. This means that the impact of an increase in water level or discharge on the number of deaths depends on the current value of these variables. Furthermore, the interaction between water level and discharge further complicates interpretation, as it indicates that the effect of water level on the number of deaths depends on discharge, and vice versa. To understand how the water level and discharge of the Kalamu River influence the number of deaths during floods, it is essential to analyze the model equation and the McFadden R-squared value. Let's start with McFadden's R-squared, which is 0.999, but this high value could be misleading and indicate overfitting given the small number of observations. This value, extremely close to 1, suggests that the model explains almost all the variance in the number of deaths. In other words, the model appears to capture virtually all the information in the data regarding the relationships among water level, discharge, and the number of deaths. However, it is crucial to interpret this value with great caution, as such proximity to 1 may signal overfitting of the model to the training data. This means the model could be overly specific to the data used for its estimation and perform poorly when applied to new data.

The model equation and McFadden's R-squared value suggest that water level, discharge, and their interaction are important factors influencing the number of deaths during floods. However, it is crucial to consider the risk of overfitting and interpret the results with caution, particularly given the nonlinear and interactive nature of the relationships between variables.

Table 4 presents the exponential coefficients for the Boma flood-related death prediction model.

Table 4. Exponential Coefficients and Interpretations of Model Parameters

Parameter	Rate ratio	p-value
Intercept	0.6559	0.363
Water_Level	2.004	0.0526
Discharge	0.280	0.0486
Flow_Water_Interaction	4.930	0.0004

Let's start with the intercept, which has a rate ratio of 0.6559. This coefficient represents the log of the death rate when all explanatory variables are zero. A rate ratio less than 1 indicates that, in the absence of other variables, the model predicts fewer deaths than the reference average. This suggests that, in this hypothetical scenario, the death rate is relatively low. However, the associated p-value of 0.3629 is high and exceeds 0.05. This means that the intercept is not statistically significant, preventing us from concluding that it has a tangible impact on the number of deaths. It is therefore prudent not to give this value undue weight in the overall interpretation of the model. Regarding water level, the rate ratio is 2.0040. This means that a one-unit increase in water level is associated with an approximately 100.4% increase in the death rate. This observation suggests that water levels have a positive, potentially significant effect on the risk of death from flooding. The p-value of 0.0526, while close to significance, does not conclusively reach this threshold. This indicates a trend toward water levels affecting deaths, but the evidence is not strong enough to definitively establish this impact in the context of this study. Next, discharge displays a rate ratio of 0.2800. This indicates that a one-unit increase in discharge is associated with an approximately 72% decrease in the death rate. This relationship suggests that higher flow rates could have a protective effect, potentially by promoting better water circulation and reducing the risk of stagnant water accumulation. The p-value of 0.0486, less than 0.05, indicates that this coefficient is statistically significant. Thus, we have sufficient evidence to conclude that flow rate has a tangible impact on the number of deaths, and an increase in flow rate is associated with a significant reduction in the risk of death.

Finally, the interaction between water level and flow rate has a rate ratio of 4.9301. This high figure indicates that a one-unit increase in this interaction is associated with an approximately 393% increase in the death rate. This underscores the importance of considering the combined effect of these two variables, suggesting that their interaction plays a crucial role in the risk of death. The corresponding p-value is 0.0004, indicating that the interaction effect is highly significant. This shows a high probability that this interaction has a tangible impact on the number of deaths, thereby justifying its inclusion in the model.

The analysis of the exponential coefficients and p-values provides valuable insights into the factors influencing flood-related deaths in Boma. Water level and discharge, as well as their interaction, appear to be key determinants of the risk of death. The statistical significance of the discharge and interaction coefficients underscores the importance of careful flood management, as these factors can have significant implications for public safety.

These results highlight the importance of integrated water resource management, which considers not only water level but also discharge, to minimize risks and ensure the safety of vulnerable communities.

Information criteria, such as Akaike Information Criterion (AIC) of 37.06 and Bayesian Information Criterion (BIC) of 3.73, show that lower values would be preferable for assessing model quality. Multicollinearity analysis indicates that Water_Level has a variance inflation factor (VIF) of 1.37 (acceptable). In contrast, Flow_Rate and Flow_Water_Interaction have VIFs of 3.70 and 3.25, respectively, indicating moderate multicollinearity that may affect the stability of the estimates.

Figure 3 illustrates the display of residual plots, fitted plots, and model errors.

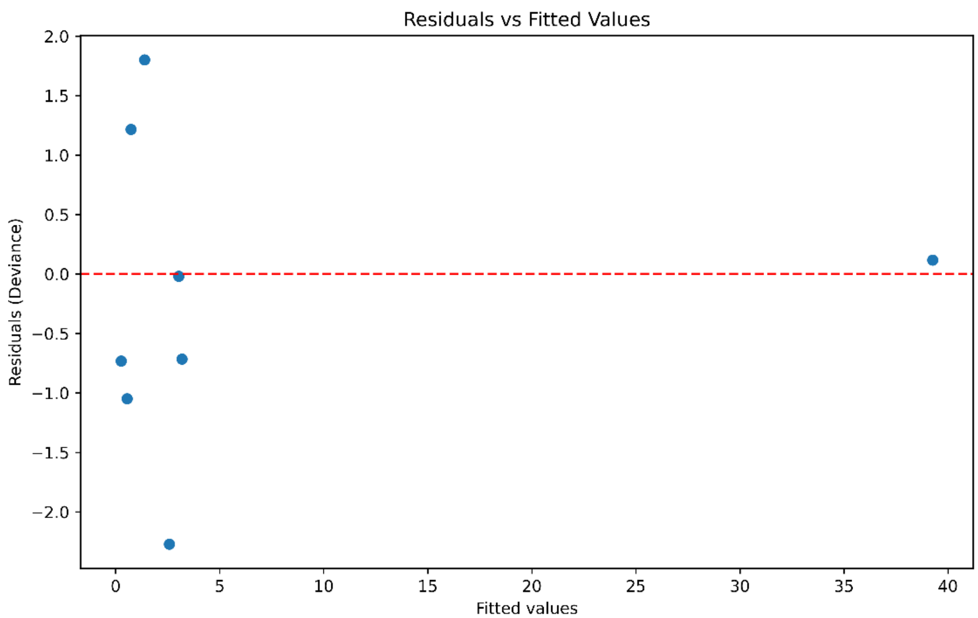

Figure 3. Residuals and Fitted Values

Figure 3 shows the relationship between residuals and predicted values. Most residuals being close to zero indicates that the model predicts relatively well for the majority of observations. However, an anomalous point that deviates significantly could signal an influential observation or a potential data problem.

Figure 4 illustrates the interaction effect via contour lines with confidence bands, relating it to hydrological processes (flow and water level).

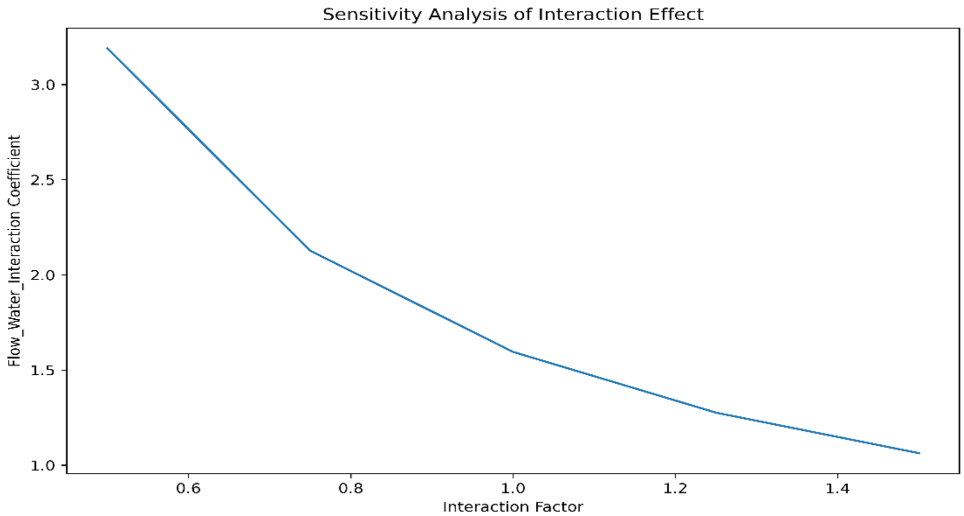

Figure 4. Sensitivity Analysis of the Interaction Effect

Figure 4 shows how the interaction coefficient between water level and discharge varies with the interaction factor. A decreasing trend is observed, suggesting that the combined effect of these variables on the number of deaths decreases as the interaction is stronger. This could indicate that at higher levels of interaction, the effects of each variable on deaths are attenuated, which is essential for risk management.

Table 5 presents the values of water level (Water_Level_Scaled), discharge (Flow_Rate_Scaled), and number of deaths (Deaths) for each observation. The Water_Level_Scaled and Flow_Rate_Scaled columns indicate that the water level and discharge values have been scaled or normalized.

Table 5. Scaled Value Set

N°	Water_Level_Scaled	Flow_Rate_Scaled	Deaths
0	-1.489	-0.530	3
1	1.221	1.133	0
2	-0.699	0.030	0
3	-0.837	-0.063	0
4	0.947	-0.111	4
5	1.599	2.323	40
6	-0.082	-1.057	0
7	0.707	-1.354	2
8	-0.357	-0.182	2
9	-1.009	-0.190	0

First, let's examine the scaled data presented in Table 5. We observe a relationship between water level, discharge, and the number of deaths. For example, when the water level and discharge are high (row 5), the number of deaths is also high (40). Conversely, when the water level and discharge are low (rows 2, 3, 6, and 9), the number of deaths is generally low (0). However, there are exceptions, such as row 1, where the water level and discharge are high but the number of deaths is low (0). This suggests that urbanization factors could also influence the number of deaths during floods.

Table 6 presents a statistical comparison of the ability of the PSO (Simplified Particle Swarm Optimization) and GA (Genetic Algorithm) algorithms to reproduce real-world data on water levels, discharge, and deaths. T-test results, including t-statistics and p-values, are presented for each variable and algorithm.

Table 6. Statistical comparison of the ability of PSO and GA algorithms to reproduce real water level, discharge, and number of deaths data (T tests)

Variable	Real data and PSO algorithm		Real data and the GA algorithm	
	T-test: t-statistic	p-value	T-test: t-statistic	p-value
Water Level Scaled	0.027	0.979	-0.771	0.451
Discharge Scaled	0.477	0.639	-0.865	0.399
Deaths	0.051	0.960	0.202	0.842

The statistical results of the T-tests show that there is no statistically significant difference between the actual data and the results simulated by the PSO and GA algorithms. This suggests that these algorithms can realistically model the relationships among water level, discharge, and the number of deaths during floods. However, it is essential to note that these statistical results do not prove that the algorithms are perfect, but rather that they can capture the main trends observed in the actual data.

When analyzing bootstrap confidence intervals for model coefficients, it is essential to note the observed variability. These confidence intervals provide an estimate of the uncertainty associated with each coefficient, which is crucial for interpreting the model's results. To begin with, the confidence interval for the intercept ranges from -40.7786 to 16.7376. This wide range indicates significant uncertainty in the intercept value, suggesting that, under certain conditions, the model's impact can vary considerably. For the coefficient associated with water level, the confidence interval is -9.9187 to 61.3772. Again, this range suggests significant variability in the effect of water level on the number of deaths. The presence of negative values indicates that, in some situations, an increase in water level may not necessarily lead to a rise in deaths, raising questions about the proper relationship between these variables. For discharge, the confidence interval ranges from -75.5325 to 48.8561. This result also highlights the uncertainty surrounding the effect of discharge on deaths. As with water levels, the presence of negative values in this interval means that increases in discharge may be associated with decreases in fatalities in some cases, but with increases in others. Finally, for the interaction term between water level and discharge, the confidence interval is -38.7170 to 59.0501. This equally wide interval reveals considerable uncertainty about how the interaction between these two variables influences the number of deaths. This suggests that the combined effect of these factors is complex and can vary considerably depending on the conditions. Overall, these wide confidence intervals indicate significant uncertainty in the coefficient estimates. This highlights the need for caution when interpreting model results, as such variability raises questions about the robustness of conclusions drawn from the current data.

Figure 5 visualizes the simulated deaths as a function of water level and flow rate from the PSO and GA algorithms above.

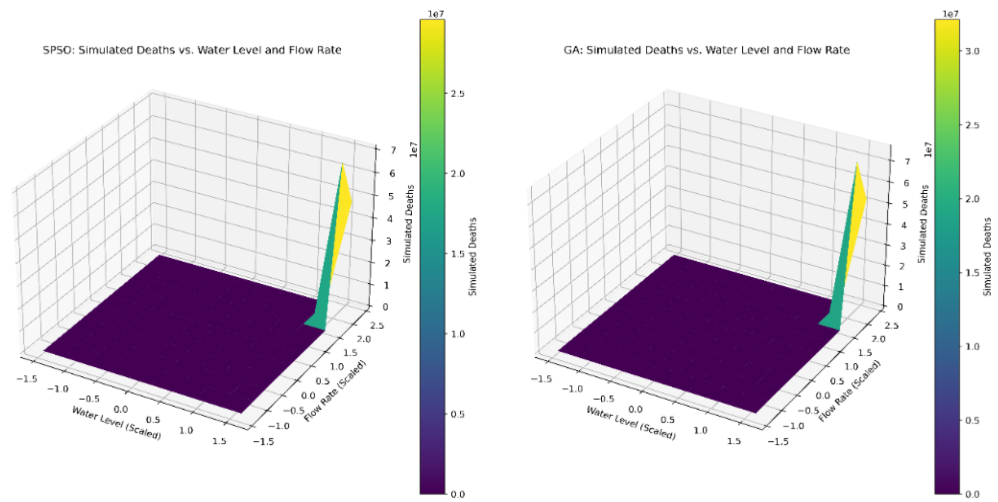

Figure 5. 3D surface of simulated deaths as a function of water level and flow rate from the PSO and GA algorithms

Based on Figure 5, which plots simulated deaths as a function of the Kalamu River's water level and discharge, several key insights can be derived about the impact of flooding. First, the figure highlights a direct relationship between water level, discharge, and the number of deaths. In other words, the higher the water level and discharge, the more simulated deaths occur. This observation supports the idea that flood intensity, as measured by these two parameters, is a determining factor in the impact on populations. Second, the figure allows us to visualize the critical thresholds where the risk of death increases significantly. By observing areas where the graph's surface rises sharply, we can identify the water levels and discharges at which the number of simulated deaths increases significantly. These thresholds can serve as reference points for flood risk planning and management, allowing alert and evacuation measures to be triggered when these levels are reached or exceeded. Furthermore, the figure compares the results from two different simulation algorithms, PSO and GA. By comparing the two graphs, the robustness of the results can be assessed, and the areas where the two algorithms converge or diverge can be identified. This comparison helps increase confidence in the conclusions drawn from the figure and better understand the uncertainties associated with the simulations. Finally, the figure can serve as a communication tool to raise public and decision-maker awareness about the impact of flooding. By presenting a clear, intuitive visualization of the relationships among water level, discharge, and the number of deaths, the figure can help mobilize resources and implement more effective flood risk prevention and management measures. This figure is a valuable tool for understanding the impact of flooding on the Kalamu River. It highlights the relationships among water level, discharge, and the number of deaths; identifies critical thresholds; and compares simulation results from different algorithms. By serving as a communication tool, it facilitates public awareness and informed decision-making. The ability of PSO and GA algorithms to realistically model the relationships among these parameters enhances their usefulness for planning and risk management.

Outre la comparaison des performances des algorithmes PSO et GA dans la modélisation prédictive des décès liés aux inondations à Boma, le résultat mentionné ci-dessus répond aux exigences d'évaluation rapportées par les chercheurs suivants.

Figure 2 compares the performance of the Simplified Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) algorithms over iterations. Both algorithms aim to minimize the fitness value, with lower values indicating better performance. PSO shows rapid improvement from the beginning, reaching a convergence of approximately -97.02, which is higher than GA's -96.51. GA, on the other hand, progresses more slowly and requires more iterations to explore the search space. These results show that PSO is more effective at quickly finding reasonable solutions. For example, Chapi et al. (2017) demonstrated that a hybrid artificial intelligence approach can improve flood risk assessment, as evidenced by the effectiveness of PSO for similar modeling problems. Furthermore, Deng et al. (2022) used hybrid approaches combining optimization algorithms for water level forecasting, highlighting the importance of convergence speed in complex settings. However, some studies, such as Di Baldassarre et al. (2013), warn of the risk of overfitting when algorithms achieve excellent performance on small training datasets. This concern may also apply to the results observed here.

Interpreting the results of the model used to predict flood-related deaths in Boma reveals several significant concerns. First, the fitness values are alarming. The root mean square error (RMSE) is 8.37, meaning that, on average, the model's predictions deviate from the actual values by about 8.37 units. This high figure indicates that the model has significant inaccuracies, making flood-related death predictions particularly problematic. Furthermore, the mean absolute error (MAE) is 6.42, reinforcing the impression of unsatisfactory performance, as this indicates that prediction errors are, on average, 6.42 units. Regarding the coefficient of determination R^2, its value is -4.04. A negative R^2 is particularly concerning because it indicates that the model fails to explain the data's variance better than a simple mean. This suggests a poor fit of the model to the data, calling for a reevaluation of the variables and modeling methods used. The results of the generalized linear model (GLM) confirm these concerns. The model, which uses a Poisson distribution, has only 10 observations, which is relatively low for robust analysis. The log-likelihood is -19.269, and the deviance is 21.522. A lower deviance value would have been desirable, as it would indicate a better fit. Also note that the pseudo-R^2 is 1.000, but this value can be misleading and may indicate overfitting, especially with so few observations. Examining the variable coefficients reveals that the intercept is -0.4218, which is not significant. For water level, the coefficient is 0.6952 ($p = 0.053$), indicating that higher water levels are associated with more deaths. On the other hand, discharge has a coefficient of -1.2731 ($p = 0.049$), suggesting that an increase in discharge is related to a decrease in deaths. The interaction term, on the other hand, has a coefficient of 1.5954 (highly significant, $p < 0.001$), highlighting that the combined effect of water level and discharge on deaths is strong. Exponential coefficient analysis reveals that the intercept (0.6559) is not significant ($p = 0.3629$). Water level (ratio of 2.0040; $p = 0.0526$) may increase the death rate. On the other hand, flow rate (0.2800, $p = 0.0486$) reduces the risk of death. The interaction between the two

(4.9301, $p = 0.0004$) is highly significant, highlighting the importance of integrated water resource management. Information criteria, such as the AIC (37.06) and BIC (3.73), indicate that lower values are preferable for assessing model quality. Analyzing multicollinearity, variance inflation factors (VIFs) suggest that water level has a VIF of 1.37 (acceptable). In contrast, discharge and the interaction term have VIFs of 3.70 and 3.25, respectively, suggesting moderate multicollinearity that may affect the stability of the estimates. T-tests show that there is no significant difference between the actual data and the results simulated by the PSO and GA algorithms, suggesting that these algorithms can realistically model the relationships between water level, discharge, and deaths. Bootstrap confidence intervals for the coefficients show significant variability. For example, the interval for the intercept is [-40.7786, 16.7376], while the interval for the water level is [-9.9187, 61.3772]. These wide intervals indicate significant uncertainty in the coefficient estimates, highlighting the need to interpret the results with caution. Several studies corroborate these results. Babatunde et al. (2015) demonstrated the importance of comparing algorithms to identify the best applications, which is relevant to PSO and GA applications. Chapi et al. (2017) highlight how hybrid approaches can improve flood-susceptibility assessment, while Deng et al. (2022) highlight the use of hybrid methods for water-level forecasting. Di Baldassarre et al. (2013) discuss human and aquatic systems relevant to flooding. Furthermore, Khosravi et al. (2019) and Hoang and Liou (2024) demonstrate the effectiveness of flood forecasting models, reinforcing the need for advanced methods to achieve accurate forecasts. The contribution of this study lies in identifying gaps in current predictive models for flood-related deaths and in proposing solutions to improve them. By integrating advanced approaches and rigorously evaluating the factors influencing forecasts, this research paves the way for significant improvements in flood risk management and the protection of vulnerable populations.

To improve the model's performance, several recommendations can be made. It would be appropriate to increase the number of observations to improve robustness. Furthermore, exploring new variables or using advanced modeling techniques, such as hybrid models combining machine learning and traditional methods, could improve the results. Finally, further validation of the models with real-world data could strengthen the reliability of flood-related death predictions.

4. Conclusions

The study presents an in-depth analysis of flood dynamics in Boma, Democratic Republic of Congo, by integrating optimization algorithms, including PSO (Simplified Particle Swarm Optimization) and Genetic Algorithms (GA), to model flood-related deaths. The results support the hypothesis that a combination of advanced methods can improve the accuracy of flood-mortality predictions by accounting for the complex interactions between water levels and river discharge. The research objectives were achieved through a rigorous methodology that included accurate data collection, data preprocessing, and the application of appropriate statistical models. The results showed that an increase in water level is directly associated with a significant increase in deaths, whereas an increase in discharge tends to reduce this risk. However, the interaction between these two variables is crucial, suggesting that high water levels with low discharge can exacerbate flood risks. Building on previous studies, this research is part of a broader approach to flood risk modeling and corroborates the findings of similar work across various geographical contexts. Studies on urban flooding have revealed comparable dynamics, in which interactions among hydrological variables play a determining role in risk predictions. The results of this study also highlight the importance of selecting optimization algorithms tailored to the problem's specific characteristics. PSO proved to be more effective than GA for this application, achieving faster convergence and better overall performance. However, it is essential to remain vigilant regarding the risks of overfitting, especially with a limited number of observations. This research offers valuable insights into flood risk management in Boma and, potentially, in other similar regions. The conclusions drawn can guide decision-makers in developing risk mitigation strategies, thereby promoting better preparedness for extreme events. The integration of statistical results and 3D visualizations enables effective communication of flood-related issues, thereby increasing awareness and community engagement in natural disaster prevention.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgements

We would like to thank Congolese Seaways and the Boma meteorological agency for their valuable support in conducting this study.

References

Babatunde, O., Armstrong, L., Leng, J. and Diepeveen, D., (2015). Comparative analysis of genetic algorithm and particle swimming optimization: an application in precision agriculture. Asian Journal of Computer and Information Systems, 3(1), 1-12.

Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., & Khosravi, K., 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environmental Modeling & Software, 95, 229–245.

Deng, B., Liu, P., Chin, R.J., Kumar, P., Jiang, C., Xiang, Y., Liu, Y., Lai, S.H., & Luo, H., 2022. Hybrid metaheuristic machine learning approach for water level prediction: A case study in Dongting Lake. Frontiers in Earth Science, 10, 928052.

Di Baldassarre, G., Kooy, M., Kemerink, J. S., Brandimarte, L., 2013. Towards understanding the dynamic behavior of floodplains as human-water systems. Hydrology and Earth System Sciences, 17(8), 3235–3244.

Fernández, D. S., Lutz, M. A., 2010. Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Engineering Geology, 111, 90–98.

Jahandideh-Tehrani, M., Bozorg-Haddad, O., Loáiciga, H.A., 2020. Application of particle swarm optimization to water management: an introduction and overview. Environ Monit Assess, 192:281

Jian, W., Li, S., Lai, C., Wang, Z., Cheng, X., Lo, E. Y. M., and Pan, T. C, 2021. Pluvial flood risk assessment in highly urbanized cities: A case study of the Pearl River Delta region of China. Natural Hazards, 105, 1691–1719.

Hoang, D.-V. and Liou, Y.-A., 2024. Elevating Flash Flood Prediction

Accurac: A Synergistic Approach with PSO and GA Optimization, Nat. Hazards Earth Syst. Sci., 1-26.

Hassan, A., Albanai, J., & Goudie, A., 2021. Modeling and Managing Flash Flood Hazards in the State of Kuwait. Preprints, 2021070011.

Kalantar, B., Ueda, N., Saeidi, V., Janizadeh, S., Shabani, F., Ahmadi, K., & Shabani, F., 2021. Deep Neural Network Using Remote Sensing Datasets for Flood Risk Susceptibility Mapping in Brisbane, Australia. Remote Sensing, 13(13), 2638

Khosravi, K., Shahabi, H., Pham, B.T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H.L., et al., 2019. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. Journal of Hydrology, 573, 311–323.

Liu, Y., Wang, L., Du, S., Zhao, L., and Liu, X., 2023. Flood forecasting method based on the improvement of VMD-FOS-QR-RBL. IEEE Access, vol. 11, pp. 4207-4218

Mudashiru, R.B., Sabtu, N., Abustan, I., & Balogun, W., 2021. Flood hazard mapping methods: A review. Journal of Hydrology, 603, 126846.

Noor, F., Laz, O.U., Haddad, K., Alim, M.A., & Rahman, A., 2022. Comparison between quantile regression technique and generalized additive model for regional flood frequency analysis: A case study for Victoria, Australia. Water, 14, 3627.

Nguyen, X. H., 2020. Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red River. Water Resources Progress, 142, 103656.

Qi, W., Ma, C., Xu, H., Chen, Z., Zhao, K., & Han, H., 2021. A review on applications of urban flood models in flood mitigation strategies. Natural Hazards, 108, 31–62.

Rentschler, J., Salhab, M., & Jafino, B.A., 2022. Flood exposure and poverty in 188 countries. Nature Communications, 13(1), 3527.

Rima, L., Haddad, K., & Rahman, A., 2025. Regional flood frequency analysis based on a generalized additive model: A parameter regression technique using the generalized extreme value distribution. Water, 17(2), 206.

Teng, J.; Jakeman, A.J.; Vaze, J.; Croke, B.F.W.; Dutta, D.; Kim, S., 2017. Flood Inundation Modeling: A Review of Methods, Recent Advances and Uncertainty Analysis. Approximately. Model. Softw, 90, 201–216.

Tien Bui, D., Pradhan, B., Nampak, H., Bui, Q.-T., Tran, Q.-A., & Nguyen, Q.-P., 2016. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. Journal of Hydrology, 540, 317–330.

Tuyen, D.N., Tuan, T.M., Son, L.H., Ngan, T.T., Giang, N.L., Thong, P.H., Hieu, V.V., Gerogiannis, V.C., Tzimos, D., & Kanavos, A., 2021. A new approach combining particle swarm optimization and deep learning for flash flood detection from satellite images. Mathematics, 9(22), 2846.

Yu, Q., Liu, C., Li, R., et al., 2023. Research on a hybrid model for flood probability prediction based on time convolutional network and particle swarm optimization algorithm. Scientific Reports, 15, 6870.

Zhen, L., & Bărbulescu, A., 2025. Quantum neural networks approach for water discharge forecasting. Applied Sciences, 15, 4119.

Zhao, G., Pang, B., Xu, Z., Peng, D., & Xu, L., 2019. Assessment of urban flood susceptibility using a semi-supervised machine learning model. Science of the Total Environment, 659, 940–94